

ABSTRACT

The present invention provides a method, system, and computer program product for extracting data from a data stream (including data streams that contain the presentation space for a legacy host screen) using a rule-based approach that does not require a user to write programming language statements. The disclosed techniques apply to presentation space data that is sent from a legacy host application to a workstation, as well as to other types of data streams (including data exchanged between applications, Web page data, etc.). Rules are defined using intuitive, interactive tools to specify the target patterns of data to be extracted. Tags in a markup language (such as the Extensible Markup Language, or "XML") are defined, and are associated with the defined rules. Upon detecting a match between the data in an incoming data stream and a target rule, an output document (expressed in the markup language) is created. Use of the markup language document provides great flexibility, enabling the document to be translated or otherwise transformed for use in multiple different environments.